

DOCUMENT RESUME

ED 220 500

TM 820 513

AUTHOR Jones, Robert F.
TITLE From MCAT to M.D.: Predicting Success in Medical School.
PUB DATE Mar 82
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (66th, New York, NY, March 19-23, 1982).
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Achievement Tests; *College Entrance Examinations; Grade Point Average; Higher Education; *Medical Schools; Medical Students; *Predictive Measurement; *Predictive Validity; Success
IDENTIFIERS *Medical College Admission Test; National Board of Medical Examiners

ABSTRACT

The effectiveness of Medical College Admission Test (MCAT) scores in predicting success during the first phase of medical education is investigated. The process by which medical students are educated and evaluated, the nature and purpose of the MCAT, and the MCAT Interpretive Studies Program developed by the Association of American Medical Colleges to study validity issues are described. Certain patterns in this preliminary data stage are that: (1) when the criteria are medical school course grades, linear composites of undergraduate college grades alone or MCAT scores alone appear to be equally predictive; (2) there is only a slight drop-off in prediction from first-year to second-year grades; (3) when the criterion is performance on National Board of Medical Examiners (NBME)-Part I, MCAT scores are substantially better predictors than undergraduate college grades; and (4) predictions of medical school course grade performance based on MCAT scores and undergraduate college grades are better than those based on either one alone; however, undergraduate college grades add little to the prediction of NBME performance by MCAT scores alone. The test scores appear to add unique information to that provided by undergraduate college grades alone, of a magnitude that is of practical significance to medical school admissions officers. (Author/PN)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

ED220500

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

* This document has been reproduced as
received from the person or organization
originating it.

Minor changes have been made to improve
reproduction quality.

Points of view or opinions stated in this docu-
ment do not necessarily represent official NIE
position or policy.

From MCAT to M.D.: Predicting Success
in Medical School

Robert F. Jones, Ph.D.
Association of American Medical Colleges

Presentation for the Symposium:
"Interpreting Data through Eternally Smudged
Glasses: The Continuing Saga of Graduate and Professional
School Validation Research"

American Educational Research Association Meeting,
New York, March, 1982

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

R. F. Jones

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

TM 820 513

In 1977, after several years of discussion, planning, and development, the New Medical College Admission Test (MCAT) was introduced to replace the old examination. The first students admitted with scores from the new test are now completing their fourth and final year of medical school, and will be awarded the M.D. degree later this spring. How well does the new test predict medical school performance? A progress report seems appropriate at this time. To provide a context for that report, I will first describe briefly 1) the process by which medical students are educated and evaluated, 2) the nature and purposes of the New MCAT, and 3) the MCAT Interpretive Studies Program, which the Association of American Medical Colleges (AAMC) has developed to study validity issues.

Figure 1 shows in somewhat oversimplified fashion the typical medical education sequence, which usually covers a period of between 6-9 years. The four years of undergraduate medical education culminate with the awarding of the M.D. degree. These years are usually divided into basic science and clinical science phases, although many schools make efforts to provide overlap by introducing students to clinical experiences and transition-type courses early in the curriculum. Course grades represent one of two major types of performance measures available during the first two years. The other is a broadly targeted test of basic science knowledge developed by an independent agency, the National Board of Medical Examiners (NBME), and known as NBME-Part I. A majority of schools require their students to take this test at the end of the second year. Course grades and NBME scores are both relevant and important criteria for

Figure 1

Medical Education Sequence

<u>Phase</u>	<u>Time in Years</u>	<u>Description</u>	<u>Major Areas</u>	<u>Performance Measures Available</u>
I basic sciences	2	lectures/laboratory in basic medical sciences	anatomy, physiology, biochemistry, pathology, microbiology, pharmacology, behavioral science	course grades, NBME-Part I scores
II clinical sciences	2	clinical clerkships; course electives	medicine, pediatrics, ob-gyn, surgery, psychiatry, family practice	clerkship grades, performance ratings, NBME-Part II scores
III graduate medical education	2-5	residency in specialty		performance ratings, NBME-Part III scores, FLEX scores

predictive validity studies, although for different reasons. The former reflects the achievement standards that each school demands of its students. The latter represents a first standard of professional competence demanded by the society at large. Passing all three Parts of the National Board exams is one of two routes to licensure for graduates of U.S. medical schools.

During the third and fourth years of medical school, the focus shifts more deliberately outside the classroom to work in hospital wards and clinics under the supervision of clinical faculty, staff physicians, and medical residents. Clerkship grades and scores on NBME-Part II, the clinical portion of the National Board sequence, are supplemented by clinical performance ratings in evaluating students during this phase. During the residency period, Part III of the National Boards is usually taken to fulfill licensure requirements. Alternatively, some students are licensed by passing the Federated Licensing Examination (FLEX), a comprehensive test which covers both basic and clinical science areas. It is sponsored by the Federation of State Medical Boards.

It is apparent from this brief description that success in medical school, even defined from a cognitive skill perspective, is multidimensional. The New MCAT was developed with this in mind. Scores are reported in six areas: Biology, Chemistry, Physics, Science Problems, Skills Analysis: Reading and Skills Analysis: Quantitative. The first three are based on a combination of knowledge and problem-solving items within disciplines, while the Science Problems score is derived from the same problem-solving items collected across disciplines. SA:Reading and SA:Quantitative comprise separate sections of the test. The thinking

skills that are the basis for science problem-solving items, and the reading and quantitative items, were derived from studies conducted by the American Institutes for Research in the Behavioral Sciences (developers of the New MCAT) for the AAMC and the National Board of Medical Examiners. These studies were concerned with critical behaviors associated with effective clinical performance during internship and residency. The Reading and Quantitative sections of the test were developed to sample thinking skills suggested by these studies to be required in the clinical settings of the third and fourth years. Evidence bearing on this issue is just beginning to be examined, and is not yet available for report. However, concurrent validity studies conducted during test development showed a tendency for the skills sections of the test to relate to performance in Introduction to Clinical Medicine courses (1). Also, recent research (2) has confirmed that the new Reading and Quantitative scales measure unique skills not covered by the Verbal Ability and Quantitative Ability subtests of the old exam.

The science portion of the test also underwent a significant transformation. While these changes were not made for the express purpose of enhancing predictive validity, they deserve mention here. Separate discipline assessments replaced a general science survey test. The content surveyed within each discipline was greatly restricted. Topics include only those which medical school faculty, students, and practicing physicians have agreed are most relevant to the medical school curriculum and future medical practice. In addition, the depth of topic coverage was limited to that found in introductory-level courses in college. More explicit descriptions of this content were made available for examinees

and users of the test score information. For all scales, the new 1-15 scoring system makes fewer discriminations.

For medical schools, these changes were intended to provide a science achievement profile for each student that 1) could be matched more directly to the academic record in applicant appraisal, 2) would permit a matching of institutional resources to student strengths and weaknesses, and 3) would be amenable to establishing minimum levels of preparation necessary for success, so that selection could include appropriate consideration of other factors, especially non-cognitive characteristics. For medical school applicants, the profile aids in evaluating their undergraduate science preparation. The more delineated content facilitates self-directed study and review. The restricted depth of coverage is an effort to avoid any disadvantage for non-science majors.

To use the new scores efficiently, schools have been quick to conduct their own local validity studies, and some have already appeared in literature. In addition, the AAMC recently invited a representative sample of 30 schools, or about one-fourth of its membership, to participate in the MCAT Interpretive Studies Program. This program is currently following two classes of students from each of these schools throughout the medical education continuum previously described. The purposes of the program are to study more extensively the range of relationships between test scores and various criterion measures at the schools, the consistency of these findings over time, and the pattern of relationships at successive stages of the educational process. I would like to report now some of the data beginning to emerge from this program,

supplemented by what we have been able to glean from other published studies.

Preliminary Findings

Table 1 shows first a summary of some overall validity indices: correlation values that result from linear composites of undergraduate GPA and MCAT scores, separately and combined. The GPA composite consists of two grade-point averages, one for science courses and one for non-science courses, which are computed for schools participating in the American Medical College Application Service (AMCAS). While these and subsequent data are based on a relatively small number of samples (classes), the samples themselves tend to be of adequate size, with median N's of between 98 and 105. Certain patterns in the data evident even at this preliminary stage can be summarized as follows:

- 1) When the criteria are medical school course grades, linear composites of undergraduate college grades alone or MCAT scores alone appear to be equally predictive.
- 2) There is only a slight drop-off in prediction from first-year to second-year grades.
- 3) When the criterion is performance on NBME-Part I, MCAT scores are substantially better predictors than undergraduate college grades.
- 4) Predictions of medical school course grade performance based on MCAT scores and undergraduate college grades are better than those based on either one alone. However, undergraduate college grades add little to the prediction of NBME performance by MCAT scores alone.

Table 1
Medians and Ranges of Multiple Correlations with
Basic Science Grades and NBME-Part I Total

	<u>Classes/Schools*</u>	<u>GPA Composite†</u>	<u>MCAT Composite</u>	<u>GPA and MCAT</u>
First Year Grades	21/11	.42 (.16 - .54)	.41 (.19 - .70)	.54 (.37 - .73)
Second Year Grades	17/9	.37 (.04 - .52)	.39 (.12 - .63)	.53 (.33 - .68)
First Two Years Combined	17/9	.43 (.10 - .52)	.41 (.14 - .65)	.54 (.35 - .71)
NBME-Part I Total	14/7	.29 (.07 - .64)	.54 (.33 - .82)	.56 (.39 - .91)

*Median class sizes for the four groups range from 98 - 105
†science GPA and non-science GPA

The first point is perhaps the most tenuous at this stage. While GPA and MCAT may be equally predictive among enrolled students, they may not be among applicants. That would occur only if these two measures received equal weight in selection, a fact disconfirmed by admissions officer's self-reports (3) and empirical analyses of admissions actions (4). While the validities of both MCAT and GPA suffer from the fact that the studies are conducted with very homogeneous enrolled student samples, the latter may be weighted more heavily in admissions and would therefore suffer more. Data presented in a subsequent table support this thesis.

There is clearly a "method" effect emerging from these data, in the tendency of grades to correlate more highly with grades and test scores with test scores. The data simply emphasize how evaluations can be influenced by the way in which they are conducted. This typically is discussed in terms of a "test-taking skill" factor, although a broader perspective would include consideration of "grade-getting skills." In truth, both grades and test scores are encumbered with trappings which hinder our view of achievement.

The fact that a combination of MCAT scores and GPA effects a better prediction than either separately is the critical datum to consider. The predictive value of standardized test scores in admissions rests in part on the assumption that they provide unique information which supplements that provided by GPA alone. In these samples, MCAT makes a considerable, unique contribution to the prediction of both course grades and National Board scores. The change in the multiple R for first year grades from .42 to .54, represents a 65 percent increase in variance explained. For NBME scores, the variance explained is almost tripled. Correspondingly,

undergraduate GPA provides a unique contribution to the prediction of course grades, but virtually none to NBME scores.

Some of the variability in these correlation values is related to the homogeneity of the various class samples. In some preliminary analyses, when the bivariate correlations (transformed to a Fisher's z) between NBME scores and individual GPA and MCAT scales were plotted against the standard deviation of the predictor, a median linear correlation of .41 was found. Table 2 provides further information on how the true validities for the MCAT and undergraduate GPA, those that would apply to applicants, are underestimated by studying the more homogeneous enrolled student samples. Along with the individual validity coefficients and those corrected for restriction of range is a selection weight index for each predictor variable. It expresses as a ratio the degree to which the enrolled student sample is more homogeneous than the corresponding applicant pool for a given class/school. A higher index suggests that relatively greater weight is given to a particular variable in admissions, either directly or indirectly (through its correlation with other variables).¹ Undergraduate science GPA has the highest median index across samples; non-science GPA and MCAT Physics the lowest. This finding is encouraging, in that it shows a correspondence between the relative weight given to each measure in admissions and the observed validities. The corrected correlations reflect these differences in selection weights. Undergraduate science GPA, by itself, is estimated to correlate .53 with first year grades among applicants. Chemistry, by itself, is estimated to correlate as high as .55 with NBME-Part I.

¹ Since the comparison is with enrolled students, not accepted students, the index is affected by student selection of the school, although in an unknown way.

Table 2
Restriction of Range Effects
on GPA and MCAT Validity Coefficients

<u>Undergraduate GPA</u>	<u>Median Selection Weight Index*</u>	<u>Median Correlationst</u>			
		<u>First Year Grades uncorrected</u>	<u>corrected</u>	<u>NBME-Part I Total uncorrected</u>	<u>corrected</u>
Science	1.91	.41	.53	.28	.39
Non-Science	1.43	.21	.26	.10	.12
Total	1.61	.37	.47	.24	.37
 <u>MCAT</u>					
Biology	1.62	.27	.36	.39	.50
Chemistry	1.83	.30	.38	.43	.55
Physics	1.41	.25	.27	.28	.34
Science Problems	1.73	.31	.37	.40	.50
SA:Reading *	1.69	.21	.27	.23	.30
SA:Quantitative	1.65	.21	.27	.29	.35

*based on a ratio Σ^2 / σ^2 , where Σ is the standard deviation in the applicant pool and σ the standard deviation for enrolled students. A higher index indicates relatively greater weight given to the measure, directly or indirectly, in admissions.

†correlations for first year grades are based on 25 classes at 15 schools. Median class size is 128. Correlations with NBME are based on 21 classes at 12 schools. Median class size is 119. The larger number of samples than in Table 1 reflects the availability of bivariate correlations from published studies.

Table 3 shows the "method" effect in more detail and provides a glimpse of discriminant validity patterns for the MCAT subtests. MCAT correlations within each basic science area are consistently higher when achievement is measured by subtests of NBME-Part I. This occurs despite the fact that anatomy, physiology, and biochemistry courses are taken typically in the first year, a full year prior to sitting for Part I of the National Board exams. Undergraduate GPA tends to correlate more highly in each area with local course grades.

As expected, the science subtests of the New MCAT tend to have stronger relationships than the skills subtests with the core biomedical science areas of the curriculum. SA:Reading and SA:Quantitative scores tend to be better predictors of performance in behavioral science. Among the science subtests, Chemistry and Science Problems tend to be the best predictors across areas of the curriculum. Physics tends to be more weakly related to every area, except physiology. Among the areas of the curriculum, performance in physiology and in biochemistry tend to be the most predictable. Discriminant patterns show up in the links of Chemistry with physiology, biochemistry, and pharmacology, and Biology with anatomy and microbiology, but they are faint at this point in the analyses.

Conclusion

These data provide a preliminary insight into how effective MCAT scores have been in predicting success during the first phase of medical education. The results, although tentative, are encouraging. The New MCAT shows moderate to strong relationships with the criteria which have been examined. More importantly, the test scores appear to add unique information to that provided by undergraduate college grades alone, of a

Table 3
Median Validity Coefficients by Basic Science Area
And Method of Evaluation*

	Anatomy		Physiology		Biochemistry		Pathology		Microbiology		Pharmacology		Behavioral Science	
	Grades	NBME	Grades	NBME	Grades	NBME	Grades	NBME	Grades	NBME	Grades	NBME	Grades	NBME
<u>GPA</u>														
Science	.32	.27	.37	.20	.36	.25	.33	.23	.31	.21	.30	.26	.20	.09
Non-science	.16	.06	.19	.06	.24	.07	.20	.18	.18	.12	.17	.13	.19	.12
Total	.30	.19	.29	.14	.34	.19	.31	.25	.27	.21	.32	.21	.24	.10
<u>MCAT</u>														
Biology	.22	.28	.26	.34	.25	.28	.17	.26	.21	.32	.13	.20	.14	.32
Chemistry	.17	.33	.34	.44	.30	.40	.17	.30	.16	.30	.19	.34	.13	.25
Physics	.15	.24	.36	.41	.19	.27	.13	.12	.06	.19	.12	.18	.10	.27
Science Problems	.16	.29	.35	.45	.27	.36	.16	.26	.17	.29	.16	.28	.16	.28
SA:Reading	.09	.11	.19	.20	.14	.10	.16	.17	.15	.18	.15	.13	.28	.41
SA:Quantitative	.14	.20	.29	.32	.21	.18	.13	.13	.14	.18	.11	.18	.17	.34

*For NBME scores, the data are based on 21 classes at 11 schools, one correlation per class. For course grades, the number of correlations in each area ranges from 12 to 18 except for anatomy where separate courses in histology, neuroanatomy, gross anatomy, and embryology bring the total number of correlations to 38. The number of classes studied ranges from 10 to 16, and the number of schools from 5 to 8.

magnitude that is of practical significance to medical school admissions officers. The validity coefficients have been shown to vary by criterion subject and criterion method of evaluation, as well as the homogeneity of the sample. As more data become available, it will be possible to estimate the relative magnitude of these effects in meta-analytic studies (4).

It will be of great interest, of course, to see if the predictive efficacy of MCAT scores extends beyond the first phase of medical education to measures of clinical performance. The criterion used in studying the validity of scholastic admission tests has generally been performance in the classroom. Yet the New MCAT includes scales intended to give an indication of future clinical performance, at least to the extent to which that performance depends upon cognitive dimensions, knowledge and thinking skills. Consideration of this issue must await a subsequent report.

The studies performed thus far have necessarily been general. For MCAT scores to be used effectively, these studies must expand in depth as well as breadth. The research questions are numerous. Are the implications of test scores for performance the same across groups: men, women, blacks, whites, science majors, non-science majors? Do the implications differ by type of medical school curriculum? How effectively can the test scores be used to presage academic difficulty? Do the separate science scores possess sufficient discriminant validity to have diagnostic value for those enrolled? By working intensively with medical schools participating in the MCAT Interpretive Studies Program over the next several years, we hope to address these and other questions.

References

1. New MCAT Interpretive Manual. Washington, D.C.:AAMC, 1977.
2. Jones, R., and Thomae-Forgues, M. A Factor Comparison of Old and New MCAT Scales. Journal of Medical Education. 56:161-166, 1981.
3. Puryear, J.B., and Lewis, L. Description of the Interview Process in Selecting Students for Admission to U.S. Medical Schools. Journal of Medical Education. 56:831-885, 1981.
4. Skager, R. On the Use and Importance of Tests of Ability in Admission to Postsecondary Education. In Ability Testing: Uses, Consequences, and Controversies. A.K. Wigdor and W. R. Garner (Eds.). Washington, D.C.:National Academy Press, 1982.
5. Glass, G.V. Integrating Findings: The Meta-Analysis of Research. Review of Research in Education, 5:351-379, 1977.